Index

σ-algebra, 15
  restriction of, 33

a.s., 27
Abel sum, 472
absolutely continuous, 31
action space, 452
adapted, 20
AdaUCB, 108, 409
admissible, 369
affine hull, 427
almost surely, 27
anytime, 78
  Exp3, 303
  Exp3-IX, 148
  Exp4, 200
  MOSS, 108
  UCB, 97
arithmetic coding, 161
Assouad’s method, 175
asymptotic optimality
  k-armed lower bounds, 177–9
  k-armed upper bounds, 97, 108, 116
  best arm identification, 355–60
  linear bandits, 258–61
  partial monitoring, 448
  ranking, 351
  Thompson sampling, 415

Bachelier–Lévy formula, 110
bandits with expert advice, 196
Bayes rule, 21
Bayesian bandit environment, 377
Bayesian optimal policy, 379
  1-armed bandit, 390
  discounted bandit, 396

Bayesian regret, 50, 379
Bayesian upper confidence bound algorithm, 415
Bellman optimality equation, 457
Bennett’s inequality, 70
Bernoulli
  bandit, 47, 57–9, 100, 112–21, 137, 140, 151,
  174, 175, 375, 377, 378, 382, 385, 386,
  390–1, 395, 401, 402, 404, 417
  distribution, 47, 62, 67, 69, 163
  random variable, 26, 43, 323
Bernstein’s inequality, 70, 70, 474
  empirical, 92, 95
beta distribution, 375
bias-variance trade-off, 193
bidual, 268, 329
bits, 162
Borel σ-algebra, 16
Borel function, see Borel measurable function
Borel measurable function, 16
Borel space, 36
boundary, 275
Bregman divergence, 269–70, 287, 383
Brengman–Huber inequality, 163, 172, 180,
  186, 251, 356
Brownian motion, 109, 338
canonical bandit model, 170
k-armed stochastic, 51
Bayesian, 384
contextual, 57
infinite-armed stochastic, 53
Carathéodory’s extension theorem, 18
cardinal optimisation, 365
cascade model, 342
categorical distribution, 71, 431, 466
<table>
<thead>
<tr>
<th>Index</th>
<th>514</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catoni’s estimator, 96</td>
<td></td>
</tr>
<tr>
<td>cell decomposition, 427</td>
<td></td>
</tr>
<tr>
<td>Cesàro sum, 456, 472</td>
<td></td>
</tr>
<tr>
<td>chain rule of probability measures, 42</td>
<td></td>
</tr>
<tr>
<td>change of measure, 31</td>
<td></td>
</tr>
<tr>
<td>Chernoff bound, 113, 350</td>
<td></td>
</tr>
<tr>
<td>chi-squared distance, 167, 175</td>
<td></td>
</tr>
<tr>
<td>chi-squared distribution, 65, 209</td>
<td></td>
</tr>
<tr>
<td>click model, 341</td>
<td></td>
</tr>
<tr>
<td>closed function, 329</td>
<td></td>
</tr>
<tr>
<td>closed set, 275</td>
<td></td>
</tr>
<tr>
<td>complement, 14</td>
<td></td>
</tr>
<tr>
<td>concave, 269</td>
<td></td>
</tr>
<tr>
<td>conditional expectation, 25</td>
<td></td>
</tr>
<tr>
<td>conditional independence, 29</td>
<td></td>
</tr>
<tr>
<td>conditional probability, 21</td>
<td></td>
</tr>
<tr>
<td>conjugate pair, 374</td>
<td></td>
</tr>
<tr>
<td>conjugate prior, 374</td>
<td></td>
</tr>
<tr>
<td>consistent policy, 177, 258</td>
<td></td>
</tr>
<tr>
<td>contextual bandit, 57, 193, 336</td>
<td></td>
</tr>
<tr>
<td>adversarial linear, 309</td>
<td></td>
</tr>
<tr>
<td>stochastic, 200, 203, 205–7</td>
<td></td>
</tr>
<tr>
<td>stochastic linear, 207–11, 309</td>
<td></td>
</tr>
<tr>
<td>contextual partial monitoring, 448</td>
<td></td>
</tr>
<tr>
<td>controlled Markov environment, 475</td>
<td></td>
</tr>
<tr>
<td>convex hull, 267</td>
<td></td>
</tr>
<tr>
<td>convex optimisation, 286</td>
<td></td>
</tr>
<tr>
<td>core set, 232</td>
<td></td>
</tr>
<tr>
<td>counting measure, 31</td>
<td></td>
</tr>
<tr>
<td>covering, 220, 227</td>
<td></td>
</tr>
<tr>
<td>Cramér transform, 68</td>
<td></td>
</tr>
<tr>
<td>Cramér–Chernoff method, 62–4, 145, 222, 223</td>
<td></td>
</tr>
<tr>
<td>cumulant generating function, 62, 120</td>
<td></td>
</tr>
<tr>
<td>cumulative distribution function, 25</td>
<td></td>
</tr>
<tr>
<td>discount factor, 336</td>
<td></td>
</tr>
<tr>
<td>discounting, 336, 394, 400</td>
<td></td>
</tr>
<tr>
<td>disintegration theorem, 42, 382</td>
<td></td>
</tr>
<tr>
<td>distribution, 15</td>
<td></td>
</tr>
<tr>
<td>domain, 267</td>
<td></td>
</tr>
<tr>
<td>dominated action, 427</td>
<td></td>
</tr>
<tr>
<td>dominating measure, 31</td>
<td></td>
</tr>
<tr>
<td>doubling trick, 78, 80, 137, 249, 299</td>
<td></td>
</tr>
<tr>
<td>dual norm, 293</td>
<td></td>
</tr>
<tr>
<td>dynamic programming, 400</td>
<td></td>
</tr>
<tr>
<td>easy partial monitoring game, 426</td>
<td></td>
</tr>
<tr>
<td>effective dimension, 216</td>
<td></td>
</tr>
<tr>
<td>elimination algorithm, 78, 79, 81, 103, 181, 236, 241</td>
<td></td>
</tr>
<tr>
<td>elliptical potential lemma, 210</td>
<td></td>
</tr>
<tr>
<td>empirical risk minimisation, 201</td>
<td></td>
</tr>
<tr>
<td>entropy, 160, 161, 167</td>
<td></td>
</tr>
<tr>
<td>events, 13</td>
<td></td>
</tr>
<tr>
<td>exchangeable, 204</td>
<td></td>
</tr>
<tr>
<td>Exp3, 131, 194, 279, 281, 286, 302, 319, 331, 333, 425</td>
<td></td>
</tr>
<tr>
<td>Exp3-IX, 143, 185, 319, 338</td>
<td></td>
</tr>
<tr>
<td>Exp3.P, 149, 201, 338</td>
<td></td>
</tr>
<tr>
<td>Exp3.S, 338</td>
<td></td>
</tr>
<tr>
<td>Exp4, 197, 215, 332</td>
<td></td>
</tr>
<tr>
<td>expectation, 23</td>
<td></td>
</tr>
<tr>
<td>explore-then-commit, 75, 91, 200, 213</td>
<td></td>
</tr>
<tr>
<td>exponential family, 100, 119, 181, 182, 357, 375, 375, 415</td>
<td></td>
</tr>
<tr>
<td>exponential weighting, 131, 322</td>
<td></td>
</tr>
<tr>
<td>algorithm, 136</td>
<td></td>
</tr>
<tr>
<td>continuous, 281–3</td>
<td></td>
</tr>
<tr>
<td>extended real line, 267</td>
<td></td>
</tr>
<tr>
<td>feasible, 458</td>
<td></td>
</tr>
<tr>
<td>feature map, 206</td>
<td></td>
</tr>
<tr>
<td>feature space, 206</td>
<td></td>
</tr>
<tr>
<td>feature vector, 206</td>
<td></td>
</tr>
<tr>
<td>feedback matrix, 424</td>
<td></td>
</tr>
<tr>
<td>Fenchel dual, 68, 222, 268</td>
<td></td>
</tr>
<tr>
<td>filtered probability space, 20</td>
<td></td>
</tr>
<tr>
<td>filtration, 20</td>
<td></td>
</tr>
<tr>
<td>finite additivity, 14</td>
<td></td>
</tr>
<tr>
<td>first-order bound, 148, 299, 304</td>
<td></td>
</tr>
<tr>
<td>first-order optimality condition, 273</td>
<td></td>
</tr>
</tbody>
</table>
Fisher information, 173
fixed design, 219
fixed share, 338
follow the leader, 287, 300
follow-the-perturbed-leader, 139, 200, 322, 406
follow-the-regularised-leader, 287, 311
changing potentials, 303
Frank–Wolfe algorithm, 234, 235
Fubini’s theorem, 31
full information, 136, 338
fundamental matrix, 477

G-optimal design, 231
gain, 455
game theory, 158
Gaussian tail lower bound, 418
generalised linear bandit, 215
generalised linear model, 213
Gittins index, 337, 393
globally observable, 429
gradient descent, 288
graph Laplacian, 217

Hahn decomposition, 24
hard partial monitoring game, 426
Hardy–Littlewood, 396, 403
heavy tailed, 63
Hedge, 136
Hellinger distance, 167
Hoeffding’s inequality, 65, 115, 350
Hoeffding’s lemma, 65, 69, 114, 136, 228
Hoeffding–Azuma, 228, 467
hopeless partial monitoring game, 426
Huffman coding, 161
hypothesis space, 371

image, 445
implicitly normalised forecaster, 305
importance-weighted estimator, 129, 130, 148, 201, 296
independent events, 21
index, 393
index policy, 393
indicator function, 17
information-directed sampling, 213, 416, 418, 448
instance-dependent bound, 201
integrable, 24
interior, 275
Ionescu–Tulcea theorem, 38, 53, 378, 454
isomorphic measurable spaces, 36
Jensen’s inequality, 269
John’s ellipsoid, 283
Kearns–Saul inequality, 69
kernel, 445
kernel trick, 211
Kiefer–Wolfowitz, 232, 238, 281, 283, 295, 310
Kraft’s inequality, 165
Kullback–Leibler divergence, 160

Laplace’s method, 223
large deviation theory, 67
law, 15
law of the iterated logarithm, 93, 104, 229
law of total expectations, 28
Le Cam’s inequality, 164
Le Cam’s method, 174
learning rate, 131
adaptive, 299, 304
time-varying, 199, 288, 303
least-squares, 219
Lebesgue integral, 23
Lebesgue measure, 24
Legendre function, 271–3, 287, 322
light tailed, 63
likelihood ratio, 230
linear subspace, 445
link function, 213
Lipschitz bandit, 215
locally observable, 429
log partition function, 375
log-concave, 281
logistic function, 375
loss matrix, 424

margin, 215
Markov chain, 38–9, 453
Markov kernel, 38
Markov policy, 453
Markov process, 41
Index

Markov property, 471
Markov reward process, 393
martingale, 39
maximal end component, 482
maximal inequality, 40, 104, 225
measurable map, 16
measurable set, 15
measurable space, 15
measure, 15
median-of-means, 95
memoryless deterministic policy, 453
memoryless policy, 453
metric entropy, 227
minimax, 103
minimax optimal, 155
mirror descent, 136, 286, 322, 333
misspecified linear bandit, 238, 255, 311, 312
model, 371
MOSS, 100, 103
multi-class classification with bandit feedback, 202
multi-task bandit, 254, 319, 330
nats, 162
negative correlation, 146
neighbouring actions, 427
non-anticipating sequence, 303
non-oblivious, 136, 299
non-parametric, 47
non-singular exponential family, 375
non-stationary bandit, 55
nonstationary, 136
null set, 27

oblivious, 294, 299
oblivious adversary, 136
one-armed bandit, 4, 58, 101
Bayesian, 389–92
online gradient descent, 288
online learning, 10, 245, 286
online linear optimisation, 286
online-to-confidence set conversion, 244
open set, 275
operator, 32
optimal experimental design, 231, 363
optimal value function, 457
optimisation oracle, 200, 326
optimism bias, 92
optional stopping theorem, 40
ordinal optimisation, 365
orthogonal complement, 445
outcome space, 13

packing, 227
parameter noise, 308
parametric, 47
Pareto optimal, 158, 369
Pareto optimal action, 427
partial monitoring, 7
partially observable Markov decision process, 471
peeling device, 104
permutation, 340
Pinsker’s inequality, 113, 119, 166, 167, 293
point-locally observable, 447
Poisson distribution, 67
policy, 52
policy iteration, 472
policy schema, 380
position-based model, 341
posterior, 370, 371
potential function, 287
predictable, 20
predictable variation, 70, 71
prediction with expert advice, 136
preimage, 14
prescriptive theory, 54
prior, 369, 371
prior variance, 374
probability distribution, 15
probability kernel, 38, 371
probability measure, 15
probability space, 15
product σ-algebra, 18
product kernel, 38
product measure, 31, 53
projective, 37
push-forward, 15
quadratic variation, 148
Index

Rademacher variable, 68
Radon–Nikodym derivative, 30
random
element, 16
variable, 16
vector, 16
random table model, 53, 129
random variable, 14
ranked bandit model, 350
ranking and selection, 365
rate function, 68
reactive adversary, 136
reduction, 307, 354
regret, 5
adversarial, 127
non-stationary, 332
policy, 136
pseudo, 55
pseudo, random, 186
random, 55
stochastic, 48
tracking, 332
regret decomposition lemma, 50
regular exponential family, 375
regular version, 42, 168, 169
regularised risk minimisation, 301
regulariser, 287
reinforcement learning, 7, 79, 400, 463
relative entropy, 162–5, 270, 383
restless bandit, 337, 402
retirement policy, 58
reward-stack model, 53, 396
ridge regression, 219
right stochastic matrix, 452

semi-bandit, 317, 320–6, 350, 352
semibandit, 417
separation oracle, 282, 460
sequential halving, 362
Sherman-Morrison formula, 218
signal variance, 374
signal-to-noise ratio, 309
signed measure, 15
similarity function, 195
simple function, 23
Sion’s minimax theorem, 300, 305, 381, 416
sliding window, 336
smoothness, 206
source coding theorem, 161
span, 456
spectral bandit, 216
state space, 452
static experts, 200
stationary transition matrix, 456, 477
stochastic optimisation, 362
stochastic process, 37
stopping rule, 39
stopping time, 39
strictly convex, 268
strongly connected component, 482
sub-σ-algebra, 15
submartingale, 39, 104
suboptimality gap, 50
sufficient statistic, 382, 390, 395, 471
of exponential family, 375
supermartingale, 39, 226
supervised learning, 196
support, 32
support function, 322, 329
supporting hyperplane, 273
theorem, 275, 384
tail probability, 25
Thompson sampling, 56, 100, 404
for reinforcement learning, 474
total variation distance, 166, 167
tower rule, 28
track-and-stop algorithm, 359
transductive learning, 202
transition matrix, 456
trivial event, 34
trivial partial monitoring game, 426
UCB-V, 95
uniform exploration algorithm, 353
union bound, 63
universal constant, xvi
unnormalised negentropy, 270, 272, 276, 289,
302, 320, 333
unnormalised negentropy, 413
unstructured bandits, 178
value function, 388
vanishing discount approach, 472
Varaiya’s algorithm, 401
VC dimension, 204
von Neumann-Morgenstern theorem, 55
Wald-Bellman equation, 388
weak neighbour, 447
weak* topology, 32
worst-case regret, 155
zeroth-order stochastic optimisation, 364